High Dynamic Range Video Compression Exploiting Luminance Masking

Yang Zhang, *Student Member, IEEE,* Matteo Naccari, *Member, IEEE,* Dimitris Agrafiotis, *Member, IEEE,* Marta Mrak, *Senior Member, IEEE,* and David R. Bull, *Fellow, IEEE*

***Abstract*—The Human Visual System (HVS) exhibits non-linear sensitivity to the distortions introduced by lossy image and video coding. This effect is due to the luminance masking, contrast masking and spatial and temporal frequency masking characteristics of the HVS. This paper proposes a novel perception-based quantization to remove non visible information in High Dynamic Range (HDR) color pixels by exploiting luminance masking so that the performance of the High Efficiency Video Coding (HEVC) standard is improved for HDR content. A prole scaling based on a tone-mapping curve computed for each HDR frame is introduced. The quantization step is then perceptually tuned on a transform unit basis. The proposed method has been integrated into the HEVC reference model for the HEVC Range Extensions (HM-RExt) and its performance was assessed by measuring the bitrate reduction against the HM-RExt. The results indicate that, the proposed method achieves significant bitrate savings, up to 42.2%, with an average of 12.8%, compared to HEVC at the same quality (based on HDR-VDP-2 and subjective evaluations).**

***Index Terms*—High dynamic range video compression, HDR imaging, perception-based video compression.**

I. Introduction

CONVENTIONAL CAPTURE and display devices, which typically employ a 24-bit per pixel encoding format, can only cover a limited dynamic range and color gamut, known as 'Standard Dynamic Range (SDR)' or 'Low Dynamic Range (LDR)'. In the real-world, luminance levels and color gamut can vary a lot more than what is captured by SDR imaging. The human visual system (HVS) can adapt to a broad range of luminance levels ranging from scotopic (10^{-5} - 10 cd/m^2) to photopic (10 - 10^6 cd/m^2) conditions [1], [2]. State-of-the-art High Dynamic Range (HDR) imaging technology (capture and display) are capable of providing high levels of immersion through the use of a dynamic range that meets or exceeds that of the HVS [3].

The perceptual advantage of HDR content is greatest when the content is displayed on an HDR display. New display technologies employ spatially dimmable backlight sources which modify the brightness of an LCD panel thus extending its dynamic range and color gamut [4]. The backlight can either be dual modulation (projector) or a panel of individually controllable LEDs that enhance image contrast.

HDR content more faithfully depicts the luminance levels and color gamut present in natural scenes (typically exceeding 6 orders of magnitude) by representing them as floating point values. The increase in immersion offered by HDR content hence comes at the cost of higher bit depth and bitrate requirements - this increased bit depth and the resulting larger memory and bandwidth requirements of HDR content are factors that could limit its widespread adoption. And this in turn necessitates the development of efficient HDR-relevant compression solutions.

Several HDR image formats have been proposed. The Radiance RGBE (.hdr) [5] file format uses a run-length coding strategy to represent 32 bits-per-pixel data. The OpenEXR (.exr) [6] file format encodes pixels using 16-bit half floating point values for the R, G and B channels respectively. The LogLuv TIFF (.tiff) [7] file format encodes HDR pixels using an 8- or 16-bit logarithmic representation for the luminance channel and 8-bit for each of the two CIE 1976 chrominance channels (u', v'). All of these HDR image formats apply lossless compression to the image data in order to preserve photometric and colorimetric pixel values within the visible dynamic range and color gamut. Moreover, color spaces such as the one proposed in [8] can be used to faithfully retain the high dynamic range and using integer values to represent pixel components.

A major challenge for HDR content is thus to develop means for efficiently compressing the data without compromising perceptual quality. This requirement is evidenced by the recent investigations in MPEG on HDR and Wide Color Gamut (WCG) [9] as well as the recent definition of HEVC Range Extension (HEVC-RExt) which specifies profiles up to 16-bits [10]. Previously proposed HDR compression methods can be classified into two categories: backward-compatible HDR compression [11]–[15] and perception-based HDR compression [16]–[20]. Backward-compatible HDR compression methods maintain compatibility with legacy 8-bit decoders allowing them to decode and display tone-mapped HDR content on legacy displays. Perception-based compression methods for HDR focus on using human visual perception models to accurately capture the various masking effects experienced by the HVS under HDR conditions, so that bits are not wasted coding redundant imperceptible information.

Results from psychovisual experiments on luminance masking [21]–[23] indicate that, when a perceptually uniform transform (e.g. LogLuv) is applied, the HVS is less sensitive to coding distortions introduced in HDR image regions where the luminance is either low or high. To exploit this varying distortion sensitivity, a masking profile can be used once the luminance of the original content has been coded to *luma* values, i.e. integer values relative to a defined bit depth. By means of this profile, the average luma intensity can be related to the quantization step of each coded block [24] to remove non visible information in HDR color pixels. However with

one fixed profile being used for the whole sequence and with this profile usually remaining fixed for all input videos, no adaptation of the luminance masking phenomenon, according to the variations in luminance dynamic range, is possible.

In order to adapt luminance masking to the dynamic range characteristics of each frame, we proposed in [20] a profile scaling method based on a tone-mapping curve that is computed for each frame. Based on the profile obtained, the quantization step is perceptually tuned on a Transform Unit (TU) basis. The proposed HDR Intensity Dependent Spatial Quantization (HDR-IDSQ) tool is then integrated into a block-based motion compensated hybrid video codec according to the design we proposed in [25]. Accordingly, the inverse quantization step at the decoder is computed independently of the availability of the average luma value for each block, therefore introducing no further latency and pipe-line refactoring at the decoder. This paper builds on top of the methods we proposed in [20] and [25] and provides the following main contributions:

- Extensive evaluation of the proposed HDR-IDSQ over additional test material and coding conditions suitable for broadcasting as well as video conferencing applications.
- Evaluation of the quality provided by the proposed HDR-IDSQ by means of subjective tests according to the ITU-T BT.500 recommendation.
- Statistical evaluation of the results obtained from the subjective tests.
- Comparison with the Perceptual Quantizer (PQ) method proposed in [26].

The performance of the proposed method was assessed in terms of bitrate reductions, at similar quality levels, relative to the HEVC reference Model (HM). Quality was measured objectively using the HDR-VDP-2 metric [27] and also via subjective testing. Experimental results show that the proposed method can achieve significant bitrate savings.

The remainder of this paper is structured as follows: Section II gives an overview of the state-of-the-art in HDR video compression and presents the most relevant literature on perceptual video coding architectures. Section III reviews HVS luminance masking phenomena and describes in detail the derivation of the HDR Intensity Dependent Quantization (HDR-IDQ) profiles. Section IV describes the integration of the proposed HDR-IDQ method in the HEVC codec, while Section V presents performance results and comparisons with HEVC. Finally, Section VI concludes the paper and suggests potential future work.

II. RELATED BACKGROUND WORK

Previously proposed HDR compression approaches can be classified into two major categories [28]. Tone-mapping based backward-capable HDR compression methods (Fig. 1) are designed so that standard 8-bit codecs, that can only cope with SDR content and low dynamic range (LDR) displays, are still able to decode and display a tone-mapped version of the HDR content. HDR capable decoders are able to inverse tone-map the decoded LDR content and, by decoding the full bitstream, use the HDR-LDR residual to reconstruct the HDR content.

As shown in Fig. 1, backward-compatible HDR compression methods can be broken down into four main processing

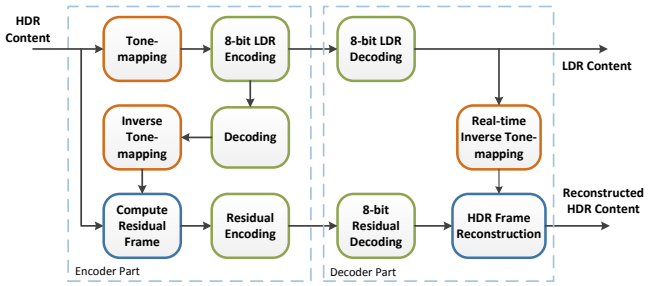


Fig. 1: General structure of backward-compatible HDR codecs.

modules. The first module contains the Tone-Mapping Operator (TMO) used for mapping HDR content down to LDR bit depths. The TMO can be adjusted in order to maximize the compression efficiency by taking into account the coding mode (i.e. intra or inter) selected by the encoder [29]. The second module calculates the HDR-LDR residual frame. This forms the side information needed for reconstructing HDR content at the decoder. This residual frame can be calculated as a ratio between the HDR and tone-mapped LDR image [30] or through a perceptually optimized nonlinear mapping function as suggested in [11]–[15]. The third module is the actual LDR coding/decoding unit that implements a method compatible with existing SDR coding architectures. The final module implements inverse tone-mapping and HDR content reconstruction through decoding of the HDR-LDR residual included in the coded bitstream.

In this type of coding strategy, the tone-mapping and inverse tone-mapping algorithms form a crucial part of the HDR codec as they can have a large impact on its rate-quality performance. Subjective assessment of various tone-mapping algorithms applied to typical HDR images and rendered in various controlled and uncontrolled environments and devices was performed in [31]. The results indicated that there is no universally best TMO for all cases, suggesting that further work on TMOs is required and that an adaptive TMO selection strategy could be beneficial.

The second approach to HDR coding [16]–[20], [26] is not backward-compatible; these methods firstly apply a perceptual quantization transformation (e.g. LogLuv in [17]) to map HDR data to the maximum bit depth supported by the encoder and then compress the input according to the selected video coding standard. There are many ways to achieve the aforementioned perceptual transformation. The work in [26] proposes a Perceptual Quantizer (PQ) to maximize the perceptual quality when creating HDR content with bit depth of 10 or 12 bits. Moreover, Zhang et al. [18] proposed a perceptual transformation which uses a Lloyd-Max (LM) quantizer for representing the luminance channel with 14 bits.

HDR-specific enhancements to a codec must clearly focus on HDR related aspects of the HVS. The HVS exhibits a non-linear sensitivity to stimuli due to luminance masking, edge masking and spatial and temporal masking phenomena. These, together with the varying HVS distortion sensitivity associated with them, can be modeled in terms of Just Noticeable Distortions (JND) [32]. JND refers to the level of distortion that can be tolerated in an image or video because it is imperceptible by a human observer. Naccari et

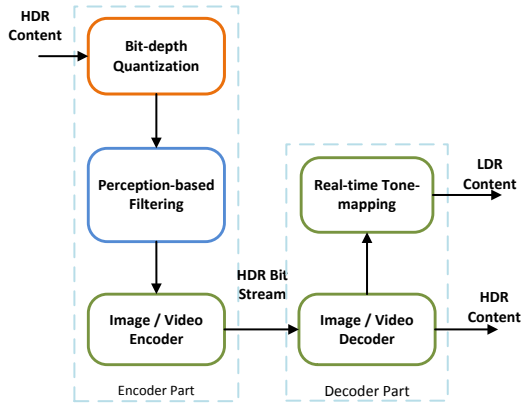


Fig. 2: Structure of the perception-based HDR codec of [35].

al. [33] proposed the use of a spatio-temporal JND model to perceptually modulate the quantization step assigned to each DCT coefficient in H.264/AVC. To avoid the encoding cost associated with communicating the varying quantization steps to the decoder, the JND model used is estimated at the decoder in [33]. The results indicate that, at the same objective quality, the proposed method provides an average bitrate reduction of about 30% compared to H.264/AVC-High profile. JND modeling has been also used in [34] to quantify the *number of distinguishable colors* provided when a particular color space is applied to HDR content. Zhang et al. proposed the application of perceptual filtering prior to HDR encoding, as shown in Fig. 2, to exploit the spatial frequency masking behavior of the HVS [35].

In [33] Naccari et al. considered the non-linear sensitivity of the HVS to background luminance variations in an image. Several psychovisual experiments have been performed to determine a luminance masking profile related to background luminance [21]–[23]. In [24] Naccari et al. propose a method for HEVC that applies coarser quantization to darker and brighter image areas. This variable quantization is performed according to a nonlinear mapping of the average luma value of an image area to the quantization step that would introduce a JND with respect to the original. The mapping from average pixel intensity to quantization step is performed through an IDQ profile that is fixed for all sequences and is a modified version of that proposed in [36].

We proposed an intensity dependent spatial quantization tool for perception-based HDR video compression in [20]. The current paper builds significantly this previous work by introducing and evaluating a low complexity and low delay realization, a method for IDQ profile transmission, a formal subjective assessment (based on an extended dataset) with statistical analysis of performance, and a comprehensive comparison of our approach with the HEVC HM.

III. PERCEPTUAL QUANTIZATION BASED ON INTENSITY MASKING

As stated in Section I, the masking phenomena exhibited by the Human Visual System can be exploited to improve HDR compression efficiency. The perceptual coding tools designed to take into account these HVS properties should also be able to adapt to the spatio-temporal nature of the coded video.

Below we present a perceptual quantization coding tool based on intensity masking and content adaptation. We first review the fundamentals of intensity masking, and then present ways of exploiting it in HDR video coding.

A. Intensity masking and the human visual system

It is well known that the human visual system shows a space and time varying sensitivity to the coding distortion introduced in images and video [32]. Intensity masking, which is exploited in the HDR coding approach described in this paper, relates to the lower sensitivity exhibited by the HVS to distortion introduced in darker and brighter image areas. Weber-Fechner's law, which states that the minimum perceivable visual stimulus difference increases with background luminance, describes this masking phenomenon. It should be noted that luminance masking is intrinsically related to the masking associated with the spatial frequency. Models to account how the distortion sensitivity varies under different luminance conditions may be complex to implement and therefore a simple multiplication of the JND profile for the luminance masking by that for spatial frequencies is usually preferred for practical reasons. In this way, the proposed HDR-IDSQ can be coupled with the JND profile for frequency masking which is commonly represented via quantization matrices, also supported in the HEVC standard.

The overall effect of intensity masking on the distortion sensitivity of the HVS is a U-shaped curve which defines the maximum amount of distortion tolerated for a given average intensity value [37]. Given that distortion in video coding is primarily introduced by quantization, intensity masking leads us to consider Intensity Dependent Quantization (IDQ). We therefore denote the associated U-shape as the IDQ profile.

In this paper, the polynomial interpolation used for this purpose is that proposed in [38] for a [0, 255] dynamic range. Considering pixel luma (μ) as representative of intensity and 8-bit pixel precision, the IDQ profile is:

$$\mathrm{IDQ}(\mu) = \begin{cases} k_1 \cdot \left(1 - \dfrac{2\mu}{256}\right)^{\lambda_1} + 1 & \mu \leq 128, \\ k_2 \cdot \left(\dfrac{2\mu}{256} - 1\right)^{\lambda_2} + 1 & \text{otherwise.} \end{cases} \quad (1)$$

The work in [38] proposes values of 2, 0.8, 3 and 2 for the parameters k_1, k_2, λ_1 and λ_2 respectively. In order to guarantee a higher content adaptation, we compute the average intensity μ over each image block B where quantization will be applied during encoding. The average pixel intensity of each block is denoted as $\mu(B)$. The IDQ profile proposed in [38] is shown in Fig. 3.

B. Intensity dependent quantization profile and video coding

The IDQ profile derived according to the procedure described in Section III-A provides an indication of the maximum amount of distortion that can be tolerated for a given intensity level. This derived IDQ profile however, does not indicate the level of quantization - the quantization step - that needs to be applied, in order to introduce the maximum

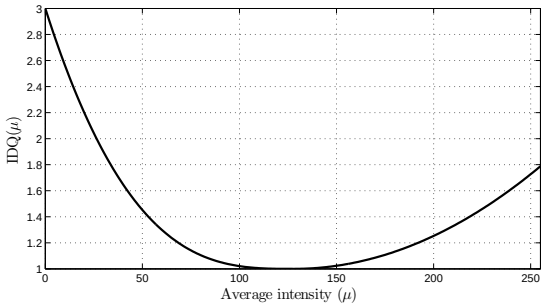


Fig. 3: The IDQ profile used in this paper as proposed in [38].

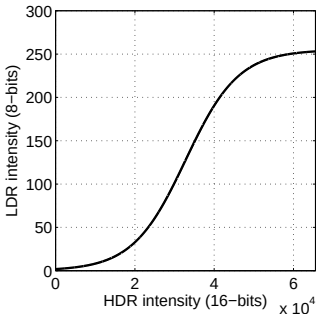


Fig. 4: Example of sigmoidal "S-shaped" tone-mapping curve.

tolerable distortion. In other words, in order to exploit an IDQ profile for improving coding efficiency there should be a link between quantization step increments and that particular IDQ profile. This link can be found by considering the quantization error (e_q) for a given frequency coefficient C in a coding block B that is quantized by a uniform quantizer with step Δ. In particular, e_q is given by:

$$e_q = C - C', \tag{2}$$

where C' denotes the reconstructed (i.e. after inverse quantization) coefficient. For a uniform quantizer, e_q is always bounded in the interval $[-\Delta/2, \Delta/2]$. From the definition of the IDQ profile, the distortion introduced by quantization will become visible if it exceeds the IDQ profile value associated to $\mu(B)$. Considering the maximum value for $|e_q|$ (i.e. $\Delta/2$), this yields:

$$\frac{\Delta}{2} \leq \text{IDQ}(\mu(B)). \tag{3}$$

Therefore, the quantization step associated with a given IDQ profile value is:

$$\Delta = 2 \cdot \text{IDQ}(\mu(B)). \tag{4}$$

Using the quantization step indicated by (4) will introduce the maximum amount of quantization noise to a coded video that will remain unnoticeable to a human observer.

The condition in (4) specifies only one rate-distortion point. To obtain different rate-distortion points, the quantization step Δ can be varied and then multiplied by $2 \cdot \text{IDQ}(\mu(B))$ to perform perceptual quantization. Hereafter, $\text{IDQ}(\mu(B))$ will simply denote the factor by which Δ is multiplied in order to exploit the intensity masking phenomenon of the HVS. In this paper, the perceptual quantization step derived in (4) is then applied to both the luma and chroma components.

C. Derivation of an HDR IDQ profile

In [23] we described an experiment on HDR luminance masking using stimuli in the range of 0.9876 to 3.5 log cd/m^2. The results indicated that the noise visibility threshold (i.e. the IDQ profile value) increases with both low and high luminance background levels and especially so with a low luminance background. In particular, the trend observed shows similarities with the IDQ profiles defined for low dynamic range imaging (e.g. that defined in Fig. 3). Therefore, an IDQ profile defined for LDR video coding can be, if properly upsampled, used for perceptual quantization in HDR coding.

There are two issues when re-using the LDR IDQ profile for HDR perceptual quantization. Firstly HDR content exhibits considerably larger brightness variations compared to LDR. Secondly, an IDQ profile derived according to the methodology described in Section III-A provides a general curve that is purely associated with the intensity masking phenomenon without taking into account the encoded content. In particular, for some images or video sequences with certain spatial andor temporal characteristics, the amount of quantization distortion introduced in a given area can be larger or smaller than the value provided by the IDQ profile.

To address these issues, we proposed a novel method in [20] for deriving the IDQ profile which is content adaptive and, at the same time, considers the larger brightness variations of HDR content. The proposed method uses nonlinear tone-mapping curves to derive the IDQ profile for HDR content. The global tone-mapping curve is a function that maps HDR luminance values to either the displayable luminance range [39], or directly to LDR pixel values with 8 bits per-channel [40]. The tone-mapping curve is usually continuous and non-decreasing with the most common shape being that of a sigmoidal ("S-shaped") function as illustrated in Fig. 4. The ultimate goal of tone-mapping is to preserve the appearance of the HDR content. Therefore a good tone-mapping operator should retain features such as intensity contrast and visual attentionfixation points. Regarding the latter, the work in [41] presents an informative study on how tone-mapping operators influence visual attention. The tone-mapping curve presented offers finer granularity where the luma histogram is denser while coarse mapping is applied to regions where the luma values are more sparse. Based on this approach and the discussion in Section III-A, the following relationship between tone-mapping curve and IDQ profile may be inferred:

- Fine tone-mapping $\rightarrow$ high sensitivity to distortion.
- Coarse tone-mapping $\rightarrow$ low sensitivity to distortion.

An HDR IDQ profile can thus be derived from an LDR one by considering the relative importance between HDR and LDR intensity values as specified by the tone-mapping curve. In the following, the above relationship is formalized through derivation of an HDR IDQ profile.

Let f denotes the tone-mapping function that maps each HDR intensity level I_{HDR} into the corresponding LDR intensity level I_{LDR}:

$$I_{\text{LDR}} = f(I_{\text{HDR}}). \tag{5}$$

Computing the differential for this relationship yields:

$$dI_{\text{LDR}} = f'(I_{\text{HDR}}) \cdot dI_{\text{HDR}}, \tag{6}$$

where f' denotes the derivate of f with respect to I_{HDR}. It should be noted that the IDQ profile (either for LDR or HDR) describes for each intensity value, the maximum amount of intensity difference that the human eye can judge as not noticeable. Therefore, by taking the differential intensities corresponding to the JND (i.e. the IDQ value), (6) becomes:

$$\text{IDQ}_{\text{LDR}} = f'(I_{\text{HDR}}) \cdot \text{IDQ}_{\text{HDR}}, \tag{7}$$

Finally the IDQ profile for HDR content is given by:

$$\text{IDQ}_{\text{HDR}} = \frac{\text{IDQ}_{\text{LDR}}}{f'(I_{\text{HDR}})}, \tag{8}$$

where IDQ_{LDR} can be any profile proposed in the literature for the LDR case.

From (8), it can be seen that, when tone-mapping is finer (i.e. high derivative values), the original values suggested by the LDR-IDQ profile are scaled or are not magnified in the HDR case. When the tone-mapping is coarse, the LDR-IDQ profile is further raised. Considering (8) it is evident that different tone-mapping curves will lead to different IDQ_{HDR} profiles. Given that tone-mapping curves are usually designed on a frame-by-frame basis, deriving IDQ_{HDR} using (8) also guarantees content adaptation.

The tone-mapping operator employed in this paper[1] is the histogram adjustment TMO proposed by Larson et al. in [40]. This uses the image histogram to implicitly segment the image so that separate scaling algorithms can be used in different areas of the luma dynamic range. Larson's TMO also employs an HVS model that takes into account aspects such as contrast and color sensitivity, glare and spatial acuity to drive the processing.

The presented formalization of the IDQ profile derivation can be also summarized as a sequence of computational steps in the following pseudo-code in Algorithm 1.

Algorithm 1: Pseudo-code for derivation of the HDR IDQ profile

```
Function: HDR IDQ profile derivation
Select an IDQ_LDR profile
Select a tone-mapping operator
for each video frame X of the HDR sequence do
    Compute the tone-mapping curve for frame X, f;
    Compute the derivate of f (f') either in closed form
    or numerically;
    for each intensity I_HDR in the HDR domain do
        Compute the LDR intensity I_LDR using (5);
        Compute the corresponding IDQ value as
        IDQ_LDR = IDQ(I_LDR);
        Compute the IDQ_HDR value using (8);
    end
end
```

[1]Note: The IDQ_{HDR} profile derivation is independent from the adopted tone-mapping curve.

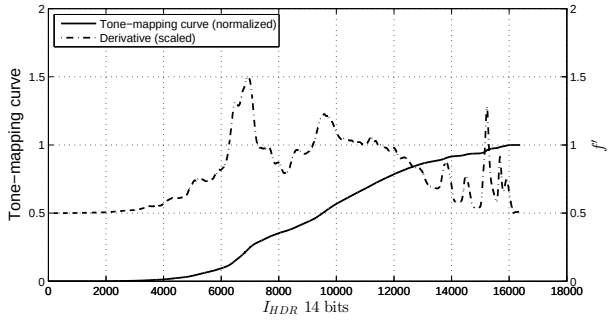


Fig. 5: Tone-mapping curve and related derivate (magnified for visibility purposes) for the first frame of the "*Library*" test sequence.

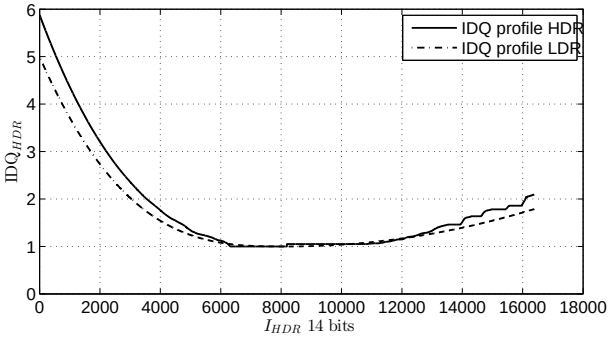


Fig. 6: IDQ_{HDR} profile derived for the first frame of the "*Library*" test sequence using the tone-mapping curve of Fig. 5.

As an example of IDQ profile derivation for HDR content, Fig. 5 shows the tone-mapping function (black continuous curve) computed on the first frame of the "*Library*" video sequence used in the tests and its corresponding derivative (dotted black line). It can be seen that where the slope of the tone-mapping curve is steep (fine mapping granularity) the associated derivative value is greater than one, which leads to an attenuated value for IDQ_{HDR} according to (8). Conversely, when the tone-mapping curve flattens (coarse mapping granularity) the IDQ_{HDR} values are magnified. In this paper, the values for f' have been rescaled in the range $[0.85\ 1.2]$ in order to tackle cases where f' tends to zero or where f' assumes large values. Such small or large f' values would lead to IDQ profile values close to zero, which might result in overflows during quantization. The values of 0.85 and 1.2 have been experimentally determined over a number of sequences.

After the tone-mapping curve and its derivative have been obtained, the IDQ_{HDR} profile can be computed using (8). Using the tone-mapping curve of Fig. 5 and the IDQ_{LDR} profile given by Fig. 3, the IDQ_{HDR} profile shown in Fig. 6 is obtained. As can been seen, the values for the dark and brighter areas are increased with respect to its LDR counterpart. In Fig. 6, the IDQ_{LDR} profile has been up-sampled to 14 bit resolution.

IV. INTEGRATING INTENSITY DEPENDENT QUANTIZATION IN THE HEVC CODEC

In this section we address the integration of IDQ in a hybrid motion-compensated predictive video codec, taking as reference the architecture of the HEVC standard. The following issues have been tackled, which arise when intensity dependent quantization is integrated in such a video codec:

1) At what level of granularity should the IDQ profile-derived quantization step be changed and how can this change be implemented with low complexity.
2) How to efficiently communicate the IDQ profile to the decoder in order to enable content adaptive perceptual quantization.
3) How to minimize the overall computational complexity cost of integrating the HDR-IDQ to the codec.

A. Quantizer integration

Based on the conclusions drawn in Section VI, exploiting the intensity masking phenomenon of the HVS leads to a modified perceptual quantization step Δ^{IDQ} given by the following equation, where Δ is the original quantization step:

$$\Delta^{\text{IDQ}} = \Delta \cdot \text{IDQ}(\mu). \tag{9}$$

According to (9) computation of the perceptual quantization step requires an additional multiplication at both encoder and decoder. In some low complexity codec implementations, the cost associated with performing this additional multiplication can be a problem; therefore other solutions to varying the quantization step may be preferable. To avoid the multiplication in (9), the relationship between the Quantization Parameter (QP) and Δ in the HEVC quantizer is considered:

$$\Delta = 2^{(\text{QP}-4)/6}. \tag{10}$$

Using (9) and inverting (10) the QP associated with Δ^{IDQ} can be computed. This perceptual QP can be expressed with respect to the base QP obtaining the intensity differential quantization parameter (idQP):

$$\begin{aligned} \text{QP}^{\text{IDQ}}(\mu) &= 6 \cdot \log_2 \underbrace{\big(\Delta \cdot \text{IDQ}(\mu)\big)}_{\Delta^{\text{IDQ}}} + 4 \\ &= \underbrace{6 \cdot \log_2(\Delta) + 4}_{\text{QP}} + 6 \cdot \log_2\big(\text{IDQ}(\mu)\big), \\ \Longrightarrow \underbrace{\text{QP}^{\text{IDQ}}(\mu) - \text{QP}}_{\text{idQP}(\mu)} &= 6 \cdot \log_2\big(\text{IDQ}(\mu)\big). \end{aligned} \tag{11}$$

In order to limit QP and idQP computed in this way to integer values, rounding is applied:

$$\text{idQP}(\mu) = \left\lfloor 6 \cdot \log_2\big(\text{IDQ}(\mu)\big) + 0.5 \right\rfloor, \tag{12}$$

By using idQP, the multiplication in (9) is replaced by a simple addition of the initial QP (associated with Δ) and the idQP (derived from the IDQ profile). It should be noted that the logarithm operation in (12) is only required for each value of the IDQ profile and not for each transform block. Moreover, the idQP values can be computed at the encoder and then transmitted to the decoder as will be explained below. In this case, the decoder does not need to compute any logarithm. The relationship between the idQP and IDQ(μ) is depicted in Fig. 7 whereby the initial IDQ profile is translated to idQP values using equation (12).

The second quantizer-related integration issue that needs to be addressed is the granularity at which the quantization step is changed. As can be seen from (12) the idQP value

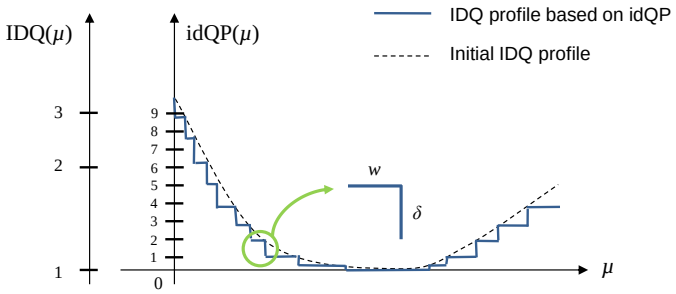


Fig. 7: idQP values relative to the corresponding IDQ profile.

depends on the average block intensity μ. This intensity refers to the original video and therefore should be transmitted to the receiver to properly decode the compressed video. This means that if the quantization step was to be changed for each coding block where frequency transform and quantization take place (a Transform Unit (TU) in HEVC), a μ value should be transmitted for each TU. This would result in an increase in the bitrate and would thus invalidate any bitrate reductions obtained through perceptual quantization. On the other hand, changing the quantization step at a coarser level of granularity (e.g. slice level) would limit content adaptation of perceptual quantization. Naccari et al. [33] proposed estimating the average intensity by using the prediction of the current TU rather than the original sample values. This average intensity estimation is also used here with the main difference that the predictor employed for intra coded TUs is the spatial one rather than the one provided by the skip mode motion vector as described in [33]. Therefore the average intensity estimate $\hat{\mu}$ is computed for each TU by averaging all the samples belonging to the spatial or temporal predictor used for the unit:

$$\hat{\mu} = \frac{1}{M \cdot N} \sum_{x=y=0}^{M,N} P(x, y), \tag{13}$$

where M, N denote the number of rows and columns forming the TU and P denotes the predictor used to compute the residual. The estimate calculated using (13) avoids the transmission of the original μ values to the decoder and allows intensity dependent quantization to be performed for each transform unit.

B. IDQ profile transmission

The profile used in intensity-dependent quantization needs to be transmitted to the decoder to ensure correct decoding of the compressed video. When using the adopted IDQ_{HDR} profile derivation, a profile needs to be transmitted, at most, on a frame basis. A straightforward profile transmission could consist of writing to the bitstream all the 2^N idQP values. For the typical bit depths used in HDR video coding, e.g. 14 bits, this would require the transmission of 16384 values.

Exploiting the staircase shape of an idQP profile (Fig. 7) can however lead to more efficient transmission strategies, reducing the bitrate associated with communicating the IDQ profile to the decoder. We propose that transmission of the IDQ profile for each frame can be achieved by first translating the profile into its idQP representation using (12). Then the obtained profile is encoded using Differential Pulse Code Modulation

(DPCM). More precisely, the whole μ range [0, 2^N-1] is divided into bins whereby each bin (b) is characterized by its width (w) and the quantity $\delta = \text{idQP}(b_m) - \text{idQP}(b_m - 1)$ as depicted in Fig. 7. The pairs (w, δ) are then written in the bitstream, more precisely in the Picture Parameter Set (PPS) syntax element [42] together with the following additional information:

- The number of (w, δ) pairs to stop the IDQ profile parsing process at the decoder.
- A flag to indicate if IDQ profile data are present.

The flag indicating the presence of profile data allows bitrate savings for static sequences where only small variations occur between subsequent frames. These frames share the same tone-mapping curve and thus will provide the same IDQ_{HDR} profile. In particular, the encoder can make an a priori analysis of intensity variations along frames and decide how many profiles should be used for the analyzed video segment.

C. Low complexity and low delay implementation

The average pixel intensity μ used for calculating the idQP value of a given coding block is estimated by computing the average intensity over all the pixels belonging to the predictor (spatial or temporal) of that coding block. At the encoder side, forward quantization uses a QP value equal to QP + idQP. At the decoder side, inverse quantization requires calculation of the value of idQP, which in turn requires the decoded pixel values of the block predictor. This dependency on the availability of the block predictor may introduce prohibitive latency in some practical decoder implementations that make use of parallel processing. For example, considering the computation of the motion compensated predictor in the decoder's pipeline, the predictor may become available only before actual reconstruction, i.e. after inverse quantization and inverse transform.

To avoid this dependency, we use the IDSQ proposed in [25]. With IDSQ, perceptual inverse quantization is performed in the spatial domain. Fig. 8 depicts the block scheme for the proposed IDSQ (the figure only considers the decoder, as the quantization step at the encoder is not dependent on the predictor availability). The entropy decoded coefficient levels in Fig. 8 can be inverse quantized to obtain the reconstructed residuals c'. The inverse transform can then be applied to obtain the perceptually scaled residuals r'. With the IDSQ tool no further latency is introduced. The perceptual (inverse) quantization is done in the "*inverse scaling*" module that provides the reconstructed residuals $\hat{r}$ which are then added back to the predictor to reconstruct the coded block. The "*inverse scaling*" module is the same as the inverse quantizer (used in the main inverse quantization stage and denoted as Q^{-1}) but operating in the spatial domain. The re-use of the same inverse quantizer avoids adding new block processing designs to the codec architecture.

V. PERFORMANCE ASSESSMENT

This section presents performance assessment results for HDR video coding using the proposed intensity dependent spatial quantization. The test material used in the assessment is first described. This is followed by a description of the coding conditions and the benchmarks against which the results were compared. The objective and subjective testing methodology and experimental setup that were used for assessing the performance are then detailed, before presenting the objective and subjective results obtained.

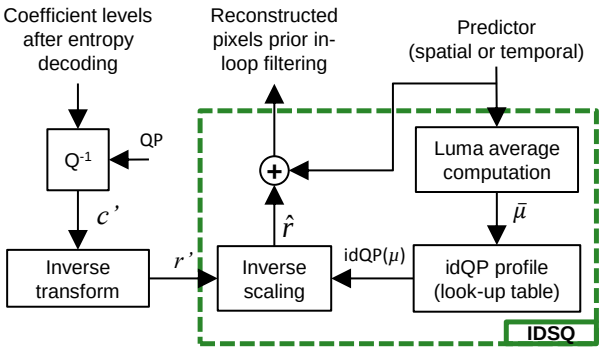


Fig. 8: Decoder block scheme for the IDSQ.

A. Test sequences

The test material used comprises five sequences from the set used by the MPEG Ad-Hoc Group (AHG) on support for HDR and XYZ color gamut [43] and a further six sequences captured using a RED EPIC camera. The MPEG AHG material consists of five sequences denotes as: "*Balloon*", "*FireEater2*", "*Market3*", "*Seine*" and "*Tibul2*". Three of these sequences, namely "*FireEater2*", "*Market3*" and "*Tibul2*", have been also selected as test material for the MPEG Call for Evidence (CfE) on HDR XYZ wide color gamut video coding [44]. The example frames from these three selected videos are shown in Fig. 9. All five clips have been captured using a rig of Sony F3 and F65 cameras at Ultra High Definition (UHD) [45] resolution (3840×2160) and at frame rates between 25 and 50 fps. The content was then down-sampled to 1920×1080 resolution and provided in OpenEXR format.

The content captured with the RED EPIC camera used the proprietary HDRx™ function which records two exposures (two video tracks) within the same interval. The primary exposure is typically normal, using the standard aperture and shutter setting. The secondary exposure is typically used for protecting the highlights in the content, and uses an adjustable shutter speed that could be 2-6 stops faster. Fig. 10 shows example frames of these test sequences, including night-time sequences ("*Carnivalx*" - (a) - (d)) and day-time sequences ("*Library*" and "*ViceChancelorRoom*" - (e) and (f) respectively). The "*Carnivalx*" sequences contain extremely bright lights coming from light bulbs and LEDs. The translational motion of the carnival cars together with the moving objects on them (dancers, rotating lamphouse etc.) represent, coding-wise, a complex scene. The "*Library*" and "*ViceChancelorRoom*" sequences contain camera panning motion. Note that the brightness level of the scene in the window is much higher than that of the indoor environment. This recorded material was captured as uncompressed RAW data with an UHD resolution and a frame rate of 50 fps. The sequences were also down-sampled to 1920×1080 resolution.

(a) **(b)** **(c)**

Fig. 9: Three example frames from the MPEG ad-hoc group test material denoted as (a) *FireEater2*, (b) *Market3*, (d) *Tibul2*. [43]

(a) **(b)** **(c)**

(d) **(e)** **(f)**

Fig. 10: Six example frames from the test HDR sequences recorded with the RED EPIC camera: Night-time sequences "*Carnivalx*": (a) *Carnival1*, (b) *Carnival2*, (c) *Carnival3*, (d) *Carnival4*. Day-time sequences: (e) "*Library*" and (f) "*ViceChancelorRoom*".

All test material was converted to the LogLuv color space with 14-bit for luminance and two 8-bit chrominance channels in 4:4:4 format. A Lloyd-Max (LM) quantizer was employed for representing the luminance channel with 14-bit [18].

B. Coding conditions

The test material was encoded using the HEVC reference test Model (HM) software. Given the 4:4:4 chroma sampling format, we employed used the Range Extensions (REext) version of HEVC. The proposed IDSQ with HDR profile derivation was also implemented with HM-10.0-REext-2.0.

The coding configurations selected for the experiments were those described in [46] using: All Intra (AI), Random Access (RA) and Low-delay B (LB) modes. The AI configuration uses only intra coding tools and it is used for low-complexity coding and fast editing, for example in the post-production phase. The RA configuration is suitable for broadcasting applications whereby a hierarchical Group Of Pictures (GOP) structure is used with an intra frame inserted depending on the video frame rate [46]. The LB configuration is representative of video conference services where low latency is required with good compression provided by inter coding tools. The aforementioned configurations cover a wide range of video coding application scenarios and allow us to extensively assess the performance of the proposed IDSQ with HDR-IDQ profile derivation. For all the tested configurations the QP values considered were 12, 17, 22 and 27.

It should be noted that, with HEVC, the minimum QP value can be reduced depending on the bit depth, according to:

$$\mathrm{QP}_{\mathrm{MIN}} = -6 \times (N - 8), \tag{14}$$

where N denotes the input bit depth in the range [8, 14[2]]. With the test material used herein having a bit depth of 14, the minimum QP can be as low as -36. Negative QP values allow the preservation of the full dynamic range of an HDR sequence as the least significant bits are not heavily quantized. However the bitrate resulting from use of such low QP values is very high and is only suitable for professional applications (e.g. studio recording) and not for broadcasting. The tested QP values have been selected as a good compromise between good perceived quality and practical bitrates.

C. Benchmarks and assessment methodology

The proposed IDSQ is compared with the following four benchmarks:

- The HM-REext codec without performing any perceptual quantization. This codec is hereafter denoted as *Anchor*.
- The HM-REext codec performing IDSQ but using a unique profile for each sequence. In this case the profile is given as the average across all frames for all profiles computed as described in Section III-C. This codec is hereafter denoted as *IDSQ-constant-profile*.
- The HM-REext codec performing IDSQ but using the profile obtained by the experiments described in [23]. This codec is hereafter denoted as *IDSQ-baseline-profile*.
- The Perceptual Quantizer (PQ) method in [26] applied to the test material which is then encoded using HM-REext. In particular, PQ is applied to obtain the same bit depth

[2]For the used HM software the maximum input bit depth supported was 14. After finalization, the REext extension specifies also 16-bit profiles.

as for the LogLuv color space and then encoded at the same bitrate of the proposed IDSQ method. This codec is hereafter denoted as *HEVC-PQ-based*[3].

The assessment methodology consists of measuring the bitrate reductions for the same video quality. These reductions are measured as percentages with respect to the *anchor*. Video quality is measured both objectively and subjectively. For the objective case, the High Dynamic Range Visible Difference Predictor (HDR-VDP-2) [27] was used. HDR-VDP-2 is a full-reference perception-based HDR image quality metric. It is acknowledged as the most recent and accurate metric for predicting the visibility of distortions affecting the luminance channel in HDR images. The metric requires the reference and the reconstructed content as well as a value for the Pixel-Per-Degree (PPD) parameter. The PPD parameter specifies the angular resolution of the image in terms of the number of pixels per visual degree and depends on the viewing distance, display resolution and display size. In our assessment the viewing distance was set to 185cm for both objective and subjective testing (a 47 inch SIM2 HDR display with a resolution of 1920×1080 pixels was employed). This resulted in a single pixel subtending a visual angle of approximately 1 *arcmin*, thus making individual pixels indistinguishable to human vision. The output of HDR-VDP-2 is a distortion map which indicates the probability of visible difference detection at each point. A higher detection probability suggests a higher distortion level at the specific point, whereas a lower detection probability suggests a lower distortion level. HDR-VDP-2 additionally incorporates a pooling strategy which provides an overall objective quality score (Q) for the test image. A non-linear logistic mapping is employed in HDR-VDP-2 to transform the objective scores Q to predicted mean opinion scores Q_{MOS}. Q_{MOS} was then used as the objective score for each video set. The performance of the proposed HDR-IDSQ tool was evaluated in terms of the bitrate reduction it offered relative to the HM codec for the same level of perceptual quality (Q_{MOS}).

As with most objective metrics, HDR-VDP-2 has limitations in terms of its correlation with subjective results. The results of [14] for example suggest that, at higher bitrates, HDR-VDP-2 scores do not increase in proportion to subjective scores. Moreover, as stated above, HDR-VDP-2 does not consider color information and therefore artifacts (e.g. chroma bleeding) associated with this channel cannot be quantified. To overcome these problems, we performed additional subjective evaluation of the proposed codec.

For subjective quality evaluation, the display was located in a dark room to minimize ambient light scatter. Participants were seated in a standardized room conforming to the ITU-R Rec. BT500-13 (Laboratory environment) [47].

The viewing distance was fixed at 185 cm, the display resolution was 1920×1080 and a 47-inch SIM2 "HDR47E" HDR display was employed. The display adopted "local dimming" technology, which has an individually controllable LED back-light array in conjunction with a Full HD resolution LCD front panel. The HDR display offers the prospect of getting much closer to the capabilities of the HVS delivering 0.4 to 4000 cd/m^2 luminance range [48].

[3]The PQ transfer function was originally proposed to grade content at 10-bit depth [26]. However, given the use of 14-bit depth in our experiments and comparisons at the same bitrate, 14 bits have also been used for the HEVC-PQ-based codec.

TABLE I: Video play out configuration with hardware and software details.

Category	Model
Processor	Intel Core i7-2600s
Graphics	Nvidia GeForce GTX 650 Ti
RAM	Kingston Memory 4GB 1333MHz
HDD (Storage)	Western Digital 500GB, 7200RPM
SSD (Playback)	Intel SSD 250GB, Read speed: 500MB/s, Write speed: 315MB/s
Operating system	Windows 7 Enterprise 64-bit
Video playback system	Matlab & Psychtoolbox-3

In the quality assessment, a set of video sequences was presented in a predefined order to a group of subjects, who were asked to rate the visual quality on a particular rating scale. Nine HDR test sequences, five from night shooting ("*Carnival2Part2*", "*Carnival4*", "*FireEater2*", "*Seine*" and "*Tibul2*") and four from day-time shooting ("*Balloon*", "*Library*", "*Market3*" and "*ViceChancelorRoom*") were selected for this study. All coding modes (AI, RA, LB) with QP = [17, 27] were evaluated in the subjective assessment and the qualities of the sequences coded by IDSQ and *HEVC-PQ-based* were compared against the *anchor*. The anchor and test sequences were displayed sequentially in randomized order for each trial. The Matlab video toolbox [49] and the PsychToolbox [50] were employed to present the HDR video sequences on the HDR display with the aid of OpenGL. The system recorded the order of the presented video sequences in each trial for data analysis. Since the typical reading speed of current Hard Disk Drives (HDD) is below 150 MB/s, a hardware solution based on Solid State Drives (SSD) was adopted for real-time HDR video playback. The details of the video play out system and the software used to display the video sequences are given in Table I.

Fig. 11 shows the flowchart of the subjective test. Since the HDR display luminance is much higher than for conventional display devices, an adaptation target image was displayed at the beginning and in-between the stimuli. The experiments started with a three-minute adaptation stage which allows the visual system of the observer to adjust to the adapting luminance. The adaptation image consists of achromatic spatial $1/f$ noise (pink noise) [51] with average luminance equal to that of each sequence and an amplitude variation of ± 0.3 log cd/m^2. The use of pink noise makes the adaptation target somewhat close to natural images in the statistical sense.

A gray circle, having an average luminance equal to that of the adaptation target, appeared in the middle of the screen. This also helped participants to stay focused on the screen during inter-stimulus intervals. Two demonstration trials took place before the formal experiment, allowing each participant to become familiarized with the task. Each trial consisted of the following steps. First, to indicate the imminence of a new trial, an acoustic signal was given one second before the display of a new HDR sequence. Then, the 1$^{\text{st}}$ sequence was displayed (full-screen) for ten seconds. An adaptation image and a notification for the 2$^{\text{nd}}$ video presentation immediately followed for two seconds. An acoustic signal was

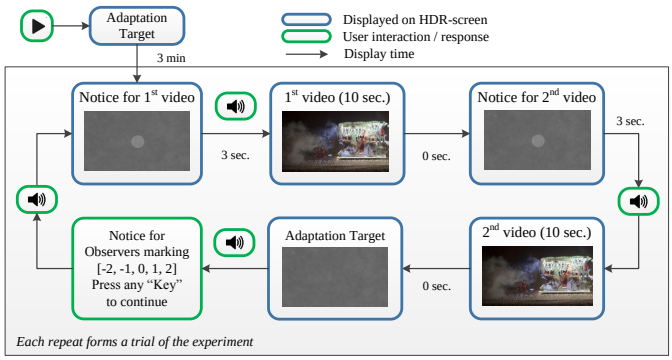


Fig. 11: Flowchart of the subjective evaluation.

then given before the display of the 2nd HDR sequence. The presentations for the anchor HDR sequence and the HDR sequence compressed by the proposed IDSQ or *HEVC-PQ-based* codecs were randomly alternated to avoid psychological conjecture. Immediately afterwards, the experiment returned to the adaptation image while awaiting the observers score. For rating the quality of the proposed method, the Stimulus-comparison method (Adjectival Categorical Judgment method) was adopted in this experiment, as recommended in BT.500-13 [47]. Participants were asked to indicate the quality of the 2nd video compared to the 1st video using a scale in the range [-2, 2] defined as -2 (Much Worse), -1 (Slightly Worse), 0 (Perceptually Equivalent), 1 (Slightly Better), 2 (Much Better). This discrete five-level scale is more suitable for naïve participants as they can quantify the quality based on noticeable differences between the two sequences. This is also more convenient for analysis compared to the case of a continuous scale.

The subjective evaluation was performed by two separate groups. In the first group, 38 participants (22 male / 16 female), ranging between 20 and 42 years of age compared the test material coded with *Anchor* and the proposed *HDR-IDSQ* codecs. In the second group, 23 participants (15 male / 8 female), ranging between 24 and 37 years of age compared the test material compressed with the *Anchor* and the *HEVC-PQ-based* codecs. The visual acuity of all participants was tested using a Snellen chart and normal color vision was tested using Ishihara tables. All participants had normal or corrected to normal visual acuity and normal color perception. All subjects were given the same instruction to ensure uniformity:

You will be shown two video sequences on the screen, one after the other, for each trial. At the end of this presentation, you will be asked whether the second video has Much Worse (-2) / Slightly Worse (-1) / Perceptually Equivalent (0) / Slightly Better (+1) / Much Better (+2) *perceptual quality than the first video. You have to choose one of the Five options i.e., [-2, 2]. Once you make a choice, press any key on the keyboard to continue to the next set of sequences.*

D. Results and discussion

To evaluate performance, the HDR-VDP-2 image quality metric was used alongside the results from subjective tests. As stated in the previous section, the HDR-VDP-2 metric provides a predicted Q_{MOS}, computed according to the probability of visible difference detection. For the experiments conducted, the predicted MOS results associated with the anchor and the proposed HDR-IDSQ tool differ only by a small amount which is considered to be below the noise level. It can therefore be concluded that the proposed IDSQ method provides perceptually equivalent quality of the anchor. The same result has been also observed over the content coded with the *IDSQ-constant-profile*, *IDSQ-baseline-profile* and *HEVC-PQ-based* codecs (with increased variability for a small number of sequences in the latter case). Given that the perceptual quality is the same, the bitrate reductions with respect to the anchor can be measured. Table II shows the bitrate values obtained by the proposed HDR-IDSQ tool, while Table III lists the reductions averaged across all tested QP points for the proposed IDSQ method as well as for the *IDSQ-constant-profile* and the *IDSQ-baseline-profile*. It can be observed that the proposed HDR-IDSQ method (IDQ profile adjusted at frame level) provides an average bitrate reduction up to 12.82% when compared to the anchor. The largest reduction (42.2%) is achieved for the "*FireEater2*" sequence in the RA coding configuration. This sequence is characterized by large and very dark image areas where the used IDQ profile can increase the quantization step to reduce the coding rate accordingly. Inspecting the data associated with the other two benchmarks, it can be observed that the savings are significantly reduced and the trade-off between using an adaptive profile and bitrate is clear. When the IDQ profile used is that obtained by subjective experiments [23], the lowest savings are obtained (*IDSQ-baseline-profile* codec). Conversely, when the IDQ profile is adaptive at sequence level, greater savings are achieved (*IDSQ-constant-profile* codec). It should was also observed that the greatest savings were obtained for high coding rates. This is expected since for low bitrates the quantization step size is large and residuals are quantized to zero regardless the IDQ adjustment.

The goal of the subjective campaign was twofold: firstly we wished to confirm that the content coded with the proposed IDSQ method had the same perceptual quality as the anchor; secondly we wanted to understand how IDSQ performance compares with video tone-mapped using PQ. The experiment involved a maximum of three participants per group, seated close to each other in three different positions (Left, Center and Right) with respect to the center of the monitor. The perceptual scores for each video set were collected and the Average (Avg) and Standard Deviation (Std) of the scores for each video set are shown in Fig. 12. As can be observed from Fig. 12a, sequences compressed with HM equipped with the proposed HDR-IDSQ tool have the same perceived quality as the anchor. Certain test content (e.g. *Library* in RA and QP 27) was actually rated higher quality than the anchor.

For the *HM-PQ-based* codec results in Fig. 12b, all results presented were at the same bit rate as the proposed HDR-IDSQ tool (see Table II). In this test, some sequences were judged to have lower quality than the anchor (e.g. *Balloon* in RA at QP 27, and *FireEater2*). Despite this the overall quality difference is marginal and the same perceptual quality assumption still holds for the PQ method.

The overall conclusion from this subjective assessment

TABLE II: The bitrate values obtained by the proposed HDR-IDSQ tool.

		AI		RA		LD	
Sequence	**QP**	**Anchor**	**HDR-IDSQ**	**Anchor**	**HDR-IDSQ**	**Anchor**	**HDR-IDSQ**
Carnival1Part1	12	361164	334982	101653	91410	123413	110813
	17	224929	208142	45218	40035	54040	47719
	22	129452	119122	18119	16022	20963	18345
	27	68833	63226	6620	5954	7038	6233
Carnival1Part2	12	350051	318931	100398	87641	121088	105703
	17	216540	197004	44593	38055	53046	45116
	22	123911	111759	17788	15110	20518	17182
	27	65426	58952	6334	5492	6752	5724
Carnival2Part1	12	320016	288137	69725	60896	86484	74568
	17	198420	177434	29977	25988	35261	30333
	22	114232	101242	12254	10733	13276	11481
	27	61336	54299	4873	4356	4684	4157
Carnival2Part2	12	340789	310974	72427	64160	90991	79476
	17	216737	196105	31849	28030	37737	32875
	22	128562	115189	13653	12143	14760	12944
	27	71827	64149	5836	5300	5673	5125
Carnival3	12	316608	294850	90698	82103	108869	97855
	17	193122	178944	40347	35957	47750	42302
	22	109141	100176	16233	14426	18517	16269
	27	57367	52333	6045	5479	6349	5667
Carnival4	12	384986	361769	137639	126174	160670	147816
	17	245874	230004	70109	63539	82438	74635
	22	146916	136328	31584	28483	37939	33871
	27	79816	73691	11949	10820	13925	12471
ViceChancelorRoom	12	124024	110396	45346	38650	55997	47225
	17	62810	55816	14151	11624	17331	14000
	22	31106	27808	3703	3146	3955	3268
	27	16478	15143	1170	1092	1017	940
Library	12	88116	82605	30688	27850	39908	35671
	17	34885	32335	6097	5269	6803	5734
	22	14992	13918	1248	1115	1122	967
	27	8225	7730	463	453	358	348
Balloon	12	85013	76597	21882	18864	23898	20423
	17	44113	39662	7251	6177	7360	5993
	22	24286	21519	2725	2421	2129	1879
	27	12044	10742	1127	1038	733	692
FireEater2	12	247189	175367	166677	112209	186873	131045
	17	150866	97816	88209	53113	103284	65034
	22	79460	43968	38685	20482	46686	25432
	27	35706	17940	13283	6725	16678	8251
Market3	12	405362	380300	71567	67167	85002	79742
	17	240138	226624	27459	26174	30952	29528
	22	134129	126200	10924	10542	10989	10657
	27	68934	65192	4483	4353	3932	3860
Seine	12	248511	244863	143329	140448	163947	160937
	17	143883	141051	68202	66422	80633	78576
	22	72040	69981	25667	24798	31821	30658
	27	31039	29851	7341	7094	8987	8637
Tibul2	12	100157	92973	35892	33285	38294	34840
	17	59757	55334	18147	16908	18057	16335
	22	32375	29656	8520	7982	7835	7142
	27	15662	14152	3551	3329	3166	2918

campaign is that the proposed IDSQ tool provides the same perceived quality as the anchor with a significantly reduced bitrate. It also important to note that since IDSQ operates inside the coding loop and is transparent to the adopted perceptual transformation, it could also be also used in combination with methods such as PQ, with the potential for further reductions in coding rate.

Additional statistical measures were computed to characterize the score distribution across the subjects in order to confirm the significance of the results. A one-sample two-tailed Students *t*-test [52] was used to determine if there existed a significant perceptual quality difference between the test video (proposed method) and the anchor video (HM codec). The null hypothesis (H_0) under test was that the scores obtained for a subject would exhibit a normal distribution with zero mean ("Perceptually equivalent") and unknown variance:

H_0: Perceptible differences $= 0$, (Perceptually Equivalent);
H_1: Perceptible differences $\neq 0$, (Statistically Significant).

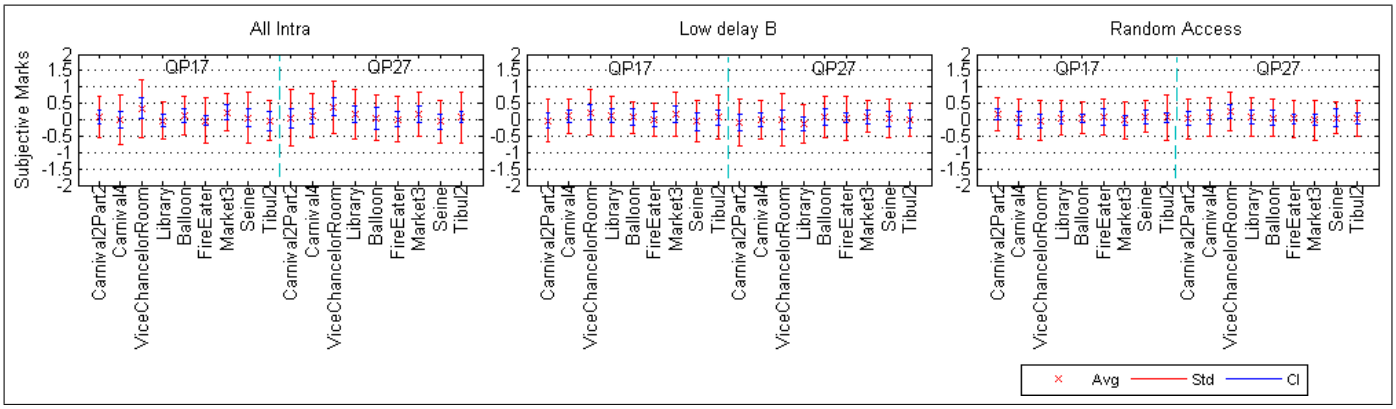
If H_0 is true, this indicates that statistically there is no difference in the perceived quality of the test content compared with the anchor sequences. On the basis of the mean score values, their standard deviations and the number of participants, the *t*-value can be computed as:

$$t = (\bar{x} - \mu)/(\sigma/\sqrt{n}), \tag{15}$$

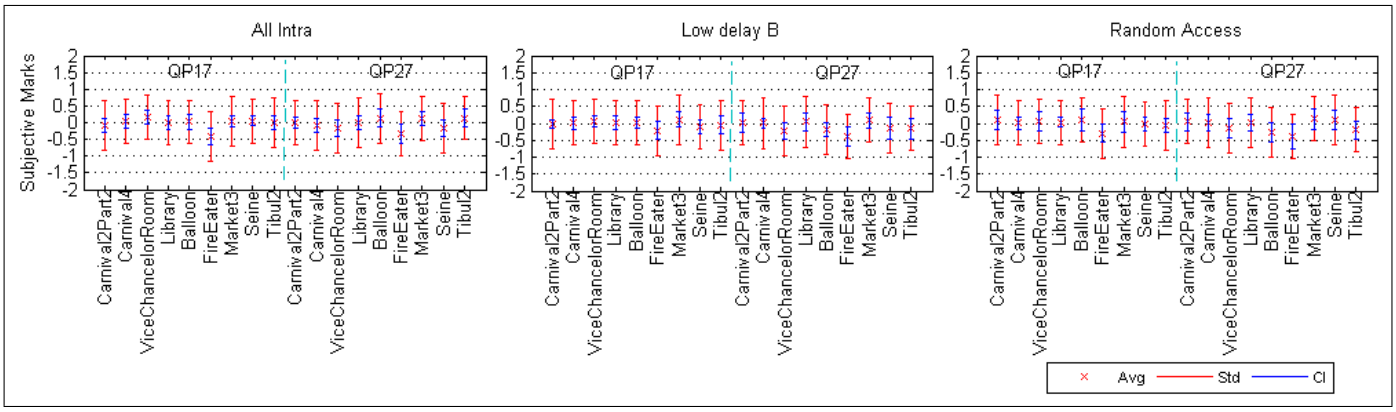
where $\bar{x}$ is the mean score for the respective cases, and μ is the population mean under the null hypothesis ($\mu = 0$ in this case), σ is the sample standard deviations for the respective cases, and n denotes the sample size (the number of observers in this case, $n = 61$). We then compared the computed *t*-value with a critical value, t_{critical}, and rejected the null hypothesis H_0 based on the condition: $t > |t_{\text{critical}}|$. Here, the absolute value of t_{critical} was used since it is a two-tailed Student *t*-test. The t_{critical} value was computed from the table for the Student *t*-distribution using a significance level $\alpha = 0.05$, which corresponds to a degree of significance of 95%. If the observed *t*-value is inside the critical range, this indicates a failure to reject the null hypothesis H_0 at the α

TABLE III: Percentage bitrate reductions for the proposed IDSQ method when the IDQ profile is frame or sequence adaptive.

	Proposed IDSQ			IDSQ-constant-profile			IDSQ-baseline-profile		
Sequence	**AI**	**RA**	**LB**	**AI**	**RA**	**LB**	**AI**	**RA**	**LD**
Carnival1Part1	-7.71	-10.79	-11.46	-1.70	-3.28	-3.53	-0.80	-0.74	-0.77
Carnival1Part2	-9.40	-13.93	-14.79	-2.50	-4.99	-5.44	-0.85	-0.92	-0.98
Carnival2Part1	-10.85	-12.25	-13.13	-2.11	-3.17	-3.51	-1.76	-1.43	-1.46
Carnival2Part2	-9.84	-10.92	-11.87	-2.14	-2.93	-3.15	-1.97	-1.69	-1.71
Carnival3	-7.80	-10.21	-11.10	-2.00	-3.26	-3.55	-1.34	-1.18	-1.27
Carnival4	-6.84	-9.24	-9.66	-2.25	-3.63	-3.78	-0.93	-0.79	-0.84
ViceChancelorRoom	-10.21	-13.57	-14.96	-2.82	-5.05	-5.66	-1.48	-1.49	-1.61
Library	-6.69	-8.91	-10.70	-0.54	-1.41	-1.88	-1.73	-1.84	-2.00
Balloon	-10.55	-11.90	-12.62	-0.60	-0.72	-0.96	-1.22	-1.03	-0.82
Fireeater	-39.66	-42.22	-40.74	-15.21	-17.33	-17.07	-4.40	-4.71	-4.89
Market3	-5.79	-4.31	-3.91	-0.02	-0.01	-0.01	-2.69	-1.41	-1.03
Seine	-2.53	-2.84	-2.98	-0.38	-0.56	-0.55	-0.23	-0.16	-0.15
Tibul	-8.15	-6.67	-8.81	-4.08	-3.44	-4.68	-0.67	-0.46	-0.53
Average	-10.46	-12.13	-12.82	-2.80	-3.83	-4.14	-1.54	-1.37	-1.39



(a)



(b)

Fig. 12: Results of the subjective test for perceptual quality comparison. (a) Comparison between anchor and the proposed HDR-IDSQ. (b) Comparison between anchor and HM-PQ-based.

significance level, meaning that the anchor and test sequences are perceptually equivalent.

As can be observed from the small confidence intervals (CI) in Fig. 12, for most test conditions the t-values are within the t_{critical} range. Hence we cannot reject the null hypothesis at the 95% significance level, meaning that the coded results based on the proposed method can be assumed to be of perceptually equivalent quality to the anchors. Three test conditions AI with QP 17, 27 and RA with QP 27 for the *ViceChancelorRoom* sequence shown in Fig. 12a, have all rejected the null hypothesis at the 95% significance level - all three test conditions indicate better perceptual scores on the test sequences. This is most likely because the *ViceChancelorRoom* sequence contains large bright and dark regions, whereas the HVS has less sensitivity to such high contrast stimuli.

In conclusion, we are confident that the proposed approach achieves better bit allocation than the HEVC HM, for content of a statistically comparable perceptual quality. Furthermore, particularly for sequences such as *ViceChancelorRoom*, the proposed method offers the potential to achieve even greater bitrate reductions.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a new intensity-dependent spatial quantization tool for perception-based HDR video

compression. The method employs a tone-mapping operator to derive the HDR-IDQ profile, which supports frame level content adaptation. Moreover, Intensity Dependent Spatial Quantization has been employed in order to minimize the latency introduced by perceptual quantization. A psychovisual study was conducted using the Stimulus-comparison method, whereby a series of reference (HM) coded vs. HDR-IDSQ coded video pairs were presented to the participants. Statistical analysis of the subjective test results indicated that there was no significant difference between the picture qualities produced by the HEVC HM and the proposed method. On this basis, it was observed that our IDSQ method offers a sizeable rate distortion performance improvement relative to the anchor - with up to 42.2% reduction in bitrate for the same quality, measured using the HDR-VDP-2 metric and confirmed through subjective testing.

In future work, we intend to address HVS sensitivities to chrominance in HDR content and explore ways of exploiting other masking phenomena.

REFERENCES

[1] K. Myszkowski, R. Mantiuk, and G. Krawczyk, *High dynamic range video*, Synthesis Lectures on Computer Graphics and Animation. Morgan & Claypool Pulishers, 2008.

[2] B. A. Wandell, *Foundations of vision*. Sinauer Associates Inc., 1995.

[3] E. Reinhard, W. Heidrich, S. Pattanaik, P. Debevec, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[4] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Transactions on Graphics (Proc. of SIGGRAPH 04)*, vol. 23, no. 3, pp. 760–768, 2004.

[5] G. J. Ward, "The RADIANCE lighting simulation and rendering system," in *Proceeding of Computer graphics and interactive techniques*. ACM, 1994, pp. 459–472.

[6] R. Bogart, F. Kainz, and D. Hess, "OpenEXR image file format," in *Proceedings of ACM SIGGRAPH, Sketches & Applications*, 2003.

[7] G. W. Larson, "LogLuv encoding for full-gamut, high-dynamic range images," *Journal of Graphics Tools*, vol. 3, no. 1, pp. 15–32, 1998.

[8] R. Mantiuk, K. Myszkowski, and H. P. Seidel, "Lossy compression of high dynamic range images and video," in *SPIE Proceedings Vol. 6057: Human Vision and Electronic Imaging XI*, 2006, p. 60570V.

[9] "Draft requirements and explorations for HDR/WCG content distribution and storage," ISO/IEC JTC1/SC29/WG11 w14510, Apr. 2014.

[10] D. Flynn, M. Naccari, C. Rosewarne, K. Sharman, J. Sole, G. Sullivan, and T. Suzuki, "High Efficiency Video Coding (HEVC) Range Extensions text specification: Draft 7," in *JCTVC-Q1005, 17th meeting, Valencia, ES*, 2014.

[11] R. Mantiuk, A. Efremov, K. Myszkowski, and H. P. Seidel, "Backward compatible high dynamic range MPEG video compression," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 713–723, 2006.

[12] A. Koz and F. Dufaux, "Methods for improving the tone mapping for backward compatible high dynamic range image and video coding," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 274–292, 2014.

[13] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, "Optimizing a tone curve for backward-compatible high dynamic range image and video compression," *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 6, pp. 1558–1571, 2011.

[14] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, "Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality," *Optical Engineering*, vol. 52, no. 10, pp. 102 008.1–102 008.15, 2013.

[15] P. Korshunov and T. Ebrahimi, "Context-dependent JPEG backward-compatible high-dynamic range image compression," *Optical Engineering*, vol. 52, no. 10, pp. 102 006.1–102 006.11, 2013.

[16] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H. P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 733–741, 2004.

[17] A. Motra and H. Thoma, "An adaptive LogLuv transform for high dynamic range video compression," in *Proceedings of IEEE International Conference Image Processing (ICIP)*, 2010, pp. 2061–2064.

[18] Y. Zhang, E. Reinhard, and D. Bull, "Perception-based high dynamic range video compression with optimal bit-depth transformation," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 1321–1324.

[19] J. U. Garbas and H. Thoma, "Temporally coherent luminance-to-luma mapping for high dynamic range video coding with H.264/AVC," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 829–832.

[20] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. Bull, "High dynamic range video compression by intensity dependent spatial quantization in HEVC," in *Proceedings of Picture Coding Symposium (PCS)*, 2013, pp. 353–356.

[21] C. Graham and E. Kemp, "Brightness discrimination as a function of the duration of the increment in intensity," *The Journal of general physiology*, vol. 21, no. 5, pp. 635–650, 1938.

[22] H. Barlow, "Temporal and spatial summation in human vision at different background intensities," *The Journal of Physiology*, vol. 141, no. 2, pp. 337–350, 1958.

[23] Y. Zhang, D. Agrafiotis, M. Naccari, M. Mrak, and D. Bull, "Visual masking phenomena with high dynamic range content," in *IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 2284–2288.

[24] M. Naccari, M. Mrak, D. Flynn, and A. Gabriellini, "Improving HEVC compression efficiency by intensity dependant spatial quantisation," in *JCTVC-J0076, 10th meeting, Stockholm, SE, July*, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), 2012.

[25] M. Naccari and M. Mrak, "Intensity dependent spatial quantization with application in HEVC," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2013, pp. 1–6.

[26] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imaging Journal*, vol. 122, no. 4, pp. 52–59, 2013.

[27] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on Graphics*, vol. 30, no. 4, p. 40, 2011.

[28] Y. Zhang, D. Agrafiotis, and D. Bull, "High dynamic range image & video compression a review," in *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–7.

[29] A. Le Dauphin, R. Boitard, D. Thoreau, Y. Olivier, E. Francois, and F. LeLéannec, "Prediction-guided quantization for video tone mapping," in *SPIE 8499, Applications of Digital Image Processing XXXVII*. International Society for Optics and Photonics, 2014, p. 92170B.

[30] G. Ward and M. Simmons, "JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG," in *Proceedings of Color Imaging Conference*, 2005, pp. 283–290.

[31] P. Korshunov and T. Ebrahimi, "A JPEG backward-compatible HDR image compression," in *Proceedings of SPIE Conference on Applications of Digital Image Processing XXXV*, 2012, p. 84990J.

[32] H. R. Wu and K. R. Rao, *Digital video image quality and perceptual coding*. CRC Press, 2005.

[33] M. Naccari and F. Pereira, "Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 766–782, 2011.

[34] T. Kunkel, G. Ward, B. Lee, and S. Daly, "HDR and wide gamut appearance-based color encoding and its quantification," in *IEEE Picture Coding Symposium*, 2013, pp. 357–360.

[35] Y. Zhang, E. Reinhard, D. Agrafiotis, and D. Bull, "Image and video compression for HDR content," in *Proceedings of SPIE Conference on Applications of Digital Image Processing XXXV*. SPIE, 2012, pp. 84 990H1–84 990H13.

[36] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 820–829, 2006.

[37] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 467–476, 1995.

[38] X. Zhang, W. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Processing*, vol. 85, no. 4, pp. 795–808, 2005.

[39] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, p. 68, 2008.

[40] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.
[41] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, "Effect of tone mapping operators on visual attention deployment," in *Proceedings of SPIE Conference on Applications of Digital Image Processing XXXV*. International Society for Optics and Photonics, 2012, pp. 1–15.
[42] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits and Systems for Video Tech*, vol. 22, no. 12, pp. 1649–1668, 2012.
[43] Y. Olivier, D. Touzé, C. Serre, S. Lasserre, F. Le Léannec, and E. François, "Description of HDR sequences proposed by Technicolor," in *m31957, 107th meeting of the Moving Picture Experts Group (MPEG)*, San José, 2014.
[44] *Call for Evidence (CfE) for HDR and WCG Video Coding*, MPEG ad-hoc group on support for HDR and XYZ, N15083, 111th meeting of the Moving Picture Experts Group (MPEG), Geneva , 2015.
[45] *Parameter values for ultra-high definition television systems for production and international programme exchange*, ITU-R Recommendation BT.2020 , Aug. 2012.
[46] D. Flynn and C. Rosewarne, "Common test conditions and software reference configurations for HEVC range extensions," in *JCTVC-L1006, 12th meeting*, Geneva, CH, 2013.
[47] *Methodology for the subjective assessment of the quality of television pictures*, ITU-Recommendation BT.500-13 , Jan. 2012.
[48] SIM2, "HDR47E S 4K specifications," 2013. [Online]. Available: http://www.sim2.com/HDR/hdrdisplay/hdr47e_s_4k
[49] D. G. Pelli, "The VideoToolbox software for visual psychophysics: Transforming numbers into movies," *Spatial vision*, vol. 10, no. 4, pp. 437–442, 1997.
[50] D. H. Brainard, "The psychophysics toolbox," *Spatial vision*, vol. 10, no. 4, pp. 433–436, 1997.
[51] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, 1987.
[52] G. W. Snedecor and W. Cochran, *Statistical methods*, 8th ed., Ames: Iowa State Univ. Press Iowa. Iowa State University Press, 1989.

Yang Zhang received his B.Eng (2008) in a joint training program between Beijing Institute of Technology, China, and the University of Central Lancashire, UK, and his M.Sc (2010) from the University of Bristol. He is currently studying for a PhD in High Dynamic Range Video Compression and expects to graduate in 2015. Yang has worked as an intern researcher at BBC Research and Development, on HEVC range extension standardization for HDR video coding. He is currently working as a research assistant in the Visual Information Laboratory in Bristol Vision Institute (BVI), University of Bristol. His research interests include high dynamic range imaging, perceptual image and video coding, human vision models, objective and subjective video quality assessment, wireless video streaming and machine learning.

Matteo Naccari was born in Como, Italy. He received the "Laurea" degree in Computer Engineering (2005) and the Ph.D. in Electrical Engineering and Computer Science (2009) from Technical University of Milan, Italy. After earning his Ph.D. he spent more than two years as a Post-doc at the Instituto de Telecomunicaes in Lisbon, Portugal affiliated with the Multimedia Signal Processing Group. Since September 2011 he joined BBC R&D as Senior Research Engineer working in the video compression team. His research interests are mainly focused in the video coding area where he works or has worked on video transcoding architectures, error resilient video coding, automatic quality monitoring in video content delivery, subjective assessment of video transmitted through noisy channels, integration of human visual system models in video coding architectures and encoding techniques to deliver Ultra High Definition (UHD) content in broadcasting applications. He also actively participates to the standardisation activities led by the Joint Collaborative Team On Video Coding (JCT-VC) where he has served as co-editor for the specification text of the HEVC Range Extensions (RExt). Matteo Naccari has authored more than 30 scientific publications on journals, conferences and book chapters.

Dimitris Agrafiotis is currently a Senior Lecturer of signal processing with the University of Bristol, Bristol, U.K. He received the M.Sc. (Distinction) degree in electronic engineering from Cardiff University, Cardiff, U.K., in 1998, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in 2002. He has worked in a number of nationally and internationally funded projects and has published more than 60 papers. His main research interests are on video coding and error resilience, video quality metrics, gaze prediction, and perceptual coding.

Marta Mrak received the Dipl. Ing. and MSc degrees in electronic engineering from the University of Zagreb, Croatia, and the PhD degree from Queen Mary University of London, UK. Before joining Research and Development Department at the BBC in 2010 to work on video compression research and the H.265/HEVC standardisation, she was a postdoctoral researcher at the University of Surrey and at Queen Mary University of London. In 2002 she was awarded a German DAAD scholarship for video compression research at Heinrich Hertz Institute, Germany. She co-authored more than 100 papers, book chapters and standardisation contributions, and co-edited a book on High-Quality Visual Experience (Springer, 2010). She has been involved in several projects funded by European and UK research councils in roles ranging from researcher to scientific coordinator. She is a member of the Multimedia Signal Processing technical committee of the IEEE, senior IEEE member, area editor of Elsevier journal Signal Processing: Image Communication and guest editor of several special issues in relevant journals in the field.

David R. Bull (M94SM07F13) received the B.Sc. degree from the University of Exeter, U.K., the M.Sc. degree from the University of Manchester, U.K., and the Ph.D. degree from the University of Wales, U.K., in 1980, 1983, and 1988, respectively. He currently holds the Chair in Signal Processing at the University of Bristol, U.K and prior to that was a Lecturer with the University of Wales, U.K., and a Systems Engineer with Rolls Royce. He was the Head of the Electrical and Electronic Engineering Department, University of Bristol, from 2001 to 2006, and is currently the Director of Bristol Vision Institute. In 2001, he co-founded ProVision Communication Technologies, Ltd., Bristol, and was its Director and Chairman until 2011. His current research interests include image and video communications and analysis for wireless, internet, military, broadcast and immersive applications. He has published some 450 academic papers and articles and has written two books on these topics and has been awarded two IEE Premiums. He also holds numerous patents, several of which have been exploited commercially. David has acted as a consultant for many major companies and organisations across the world, both on research strategy and innovative technologies. He will become an Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology from 2016.